

深层神经网络设计理念

于建国 YJango 《超智能体》作者
日本会津大学 Human Interface Lab 博士在读



CONTENT

01

智能：何谓学习

什么是智能
什么是学习
意识的作用
人类的一生

02

入门：预备知识

基础知识

03

深层：为何高效

学习的难点
无免费午餐
维度的诅咒
共享是关键

04

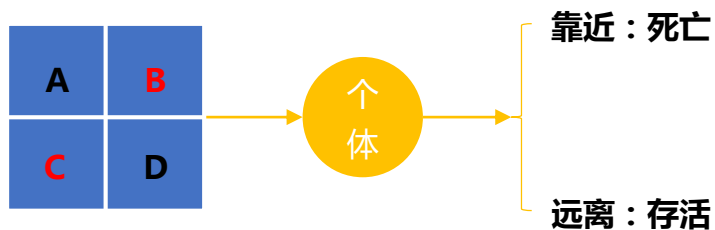
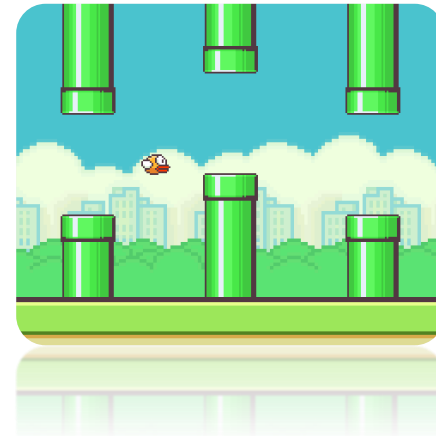
应用：设计理念

降低拟合难度
优化起点位置
增加正规惩罚
增加共享先验

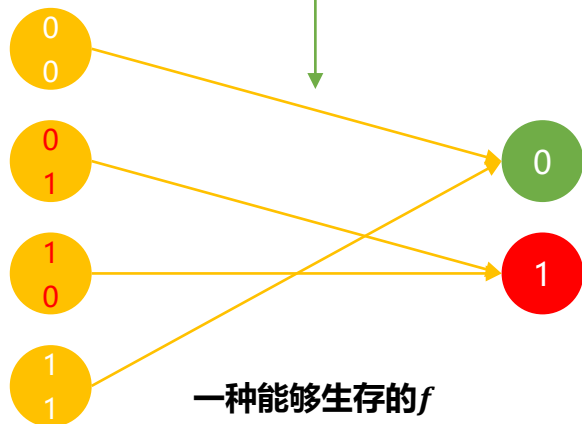
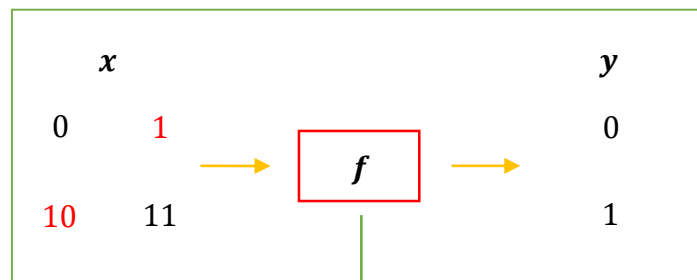
智能：何谓学习

生存游戏

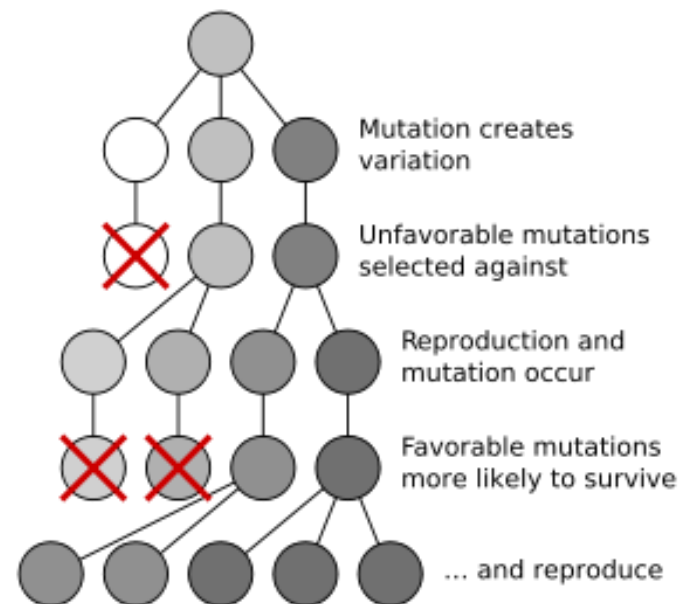
环境会变



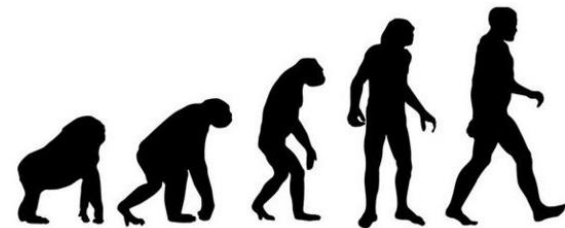
Intelligence is the ability to adapt to change



关联更新速度 > 环境变化速度



生存核心



进化是 f 的更新

进化的对象不是个体，也不是基因，而是智能
进化的过程是智能关联不断被筛选

自然选择就是学习

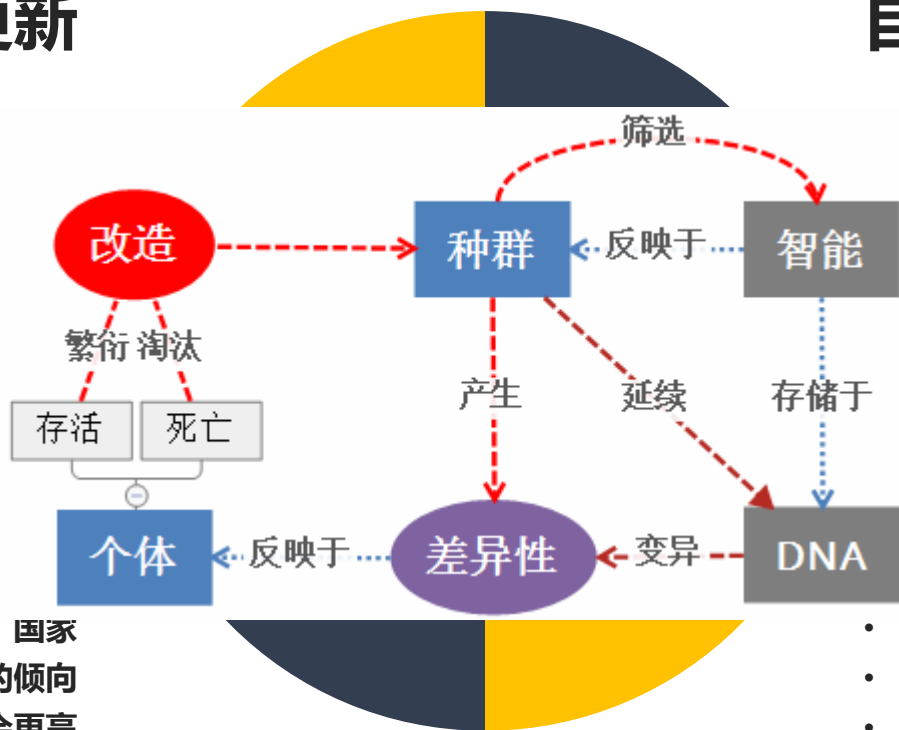
选择的动态过程就是寻找 f 的过程
它是学习的一种方式

智能的增强：选择

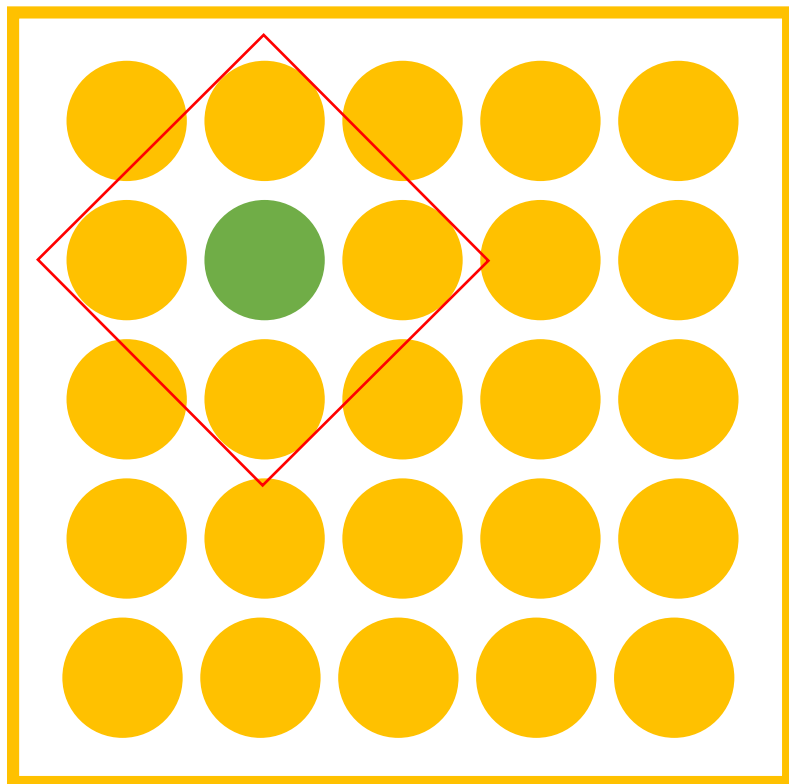
蛋白质、细胞、组织、器官、系统、个体、国家
生命有不断链接形成更高智能体从而增强智能的倾向
这样存活下去的可能性就会更高

智能是根据变化而变化的能力

- 学习： f 的寻找：自然选择
- 执行： f 的应用：蛋白质
- 延续： f 的存储：DNA



移动问题



进化的方向

进化有增加差异性的趋势。

无性繁殖到有性繁殖后，就面临着“洗牌”问题

于是生物大范围**移动**成了必然

即便是无法移动的植物，也是靠动物来帮助散播种子。

大范围移动却带来了另一个问题：

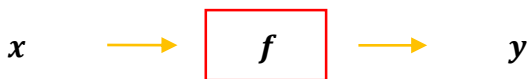
环境会随着移动而变化，上一环境所学习的 f 不再适用。



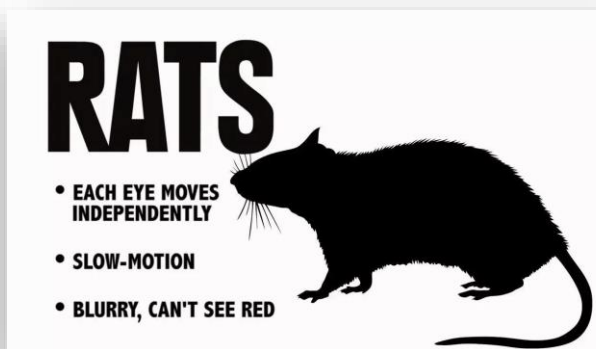
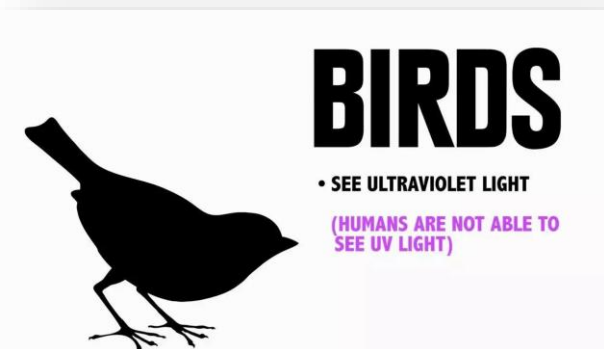
预测模型

需要新的学习方式

- 生长周期：无法自然选择试错
- 大脑模拟：预测未来事件
- 神经网络：寻找过去事件到未来事件的关联 f (任意两个空间)



视觉感知



人眼看到的并非世界的原貌，而是适合自己生存的一种感知方式。
将反光与“脑中”所看到的进行关联。
相同的反光经过不同的 f 得出不同的感知。
画面识别实际上是寻找（学习）人类的视觉关联方式 f ，并再次应用。

意识作用

智能的内容在于关联，核心在于学习，但意识还负责



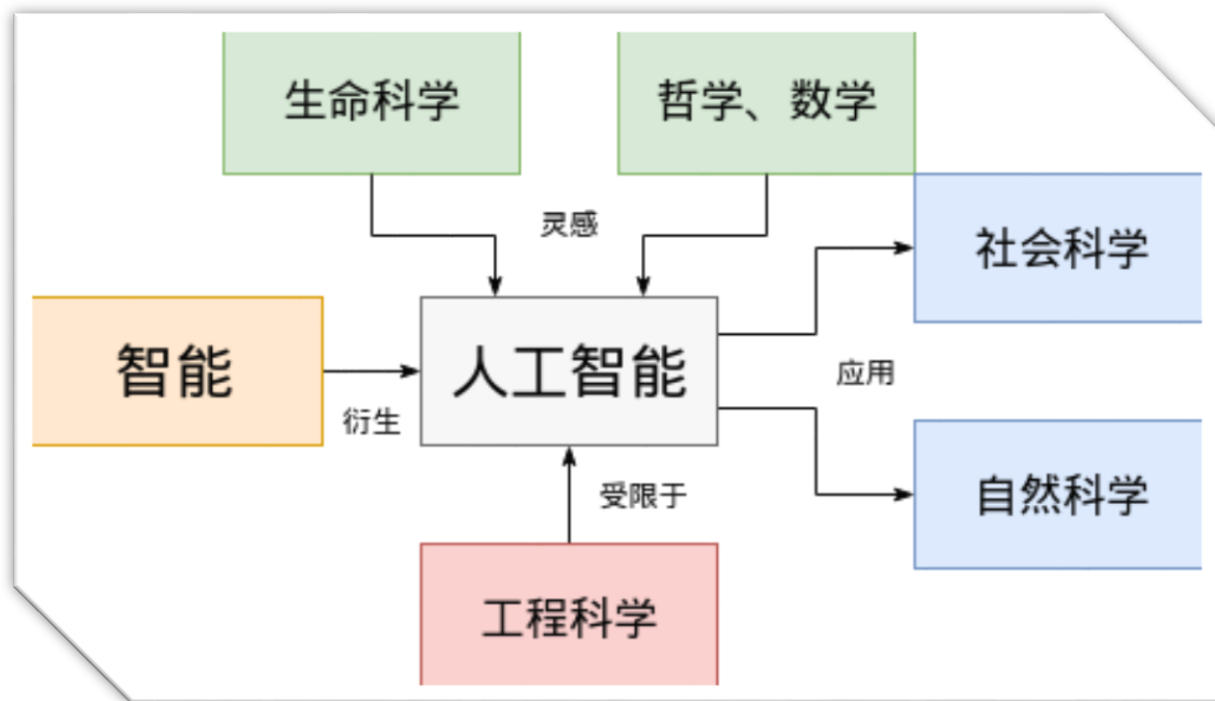
- **何时搜集数据：双向传递**
- **何时调整关联：如何说服人**
- **应用哪种关联：意识无法并行**

人的一生：建立世界模型

入门：预备知识

人工智能

在机器上实现智能



线性代数

状态与变化



概率

确信度



编程

电脑上实现



深层：为何高效

学习的难点

学习：寻找关联 f 的过程

- 任务 T ：高考数学题
- 经验 E ：历年真题
- 表现 P ：高考成绩

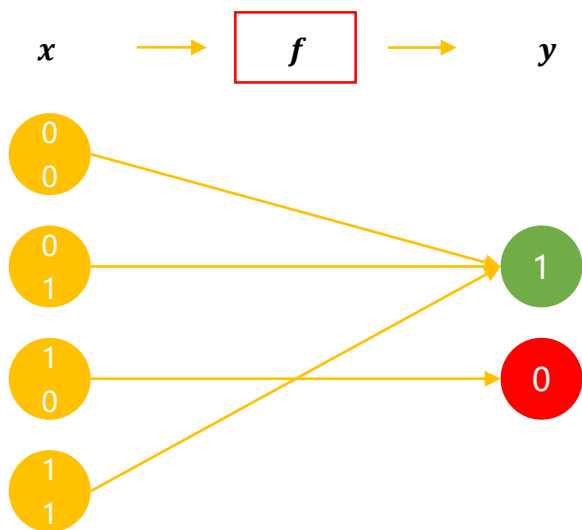


难点：需要在**未见过**的任务上表现良好



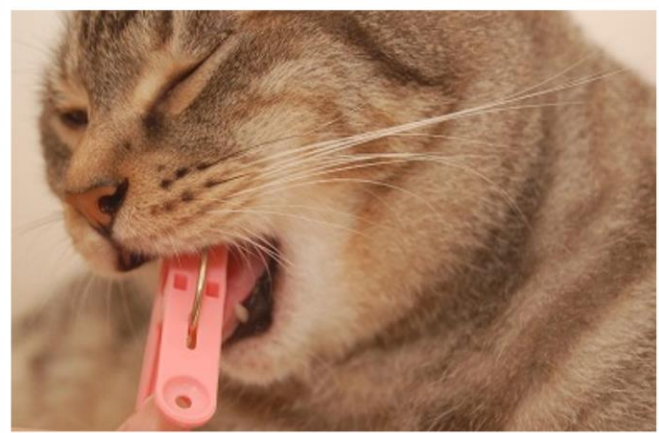
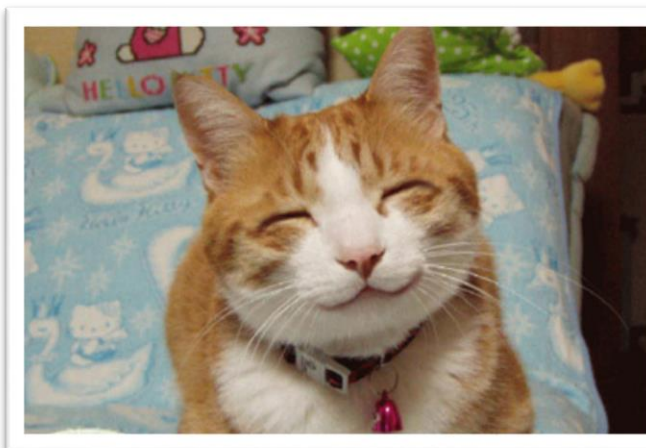
极端情况

记忆：记住**所有**的高考题和对应答案



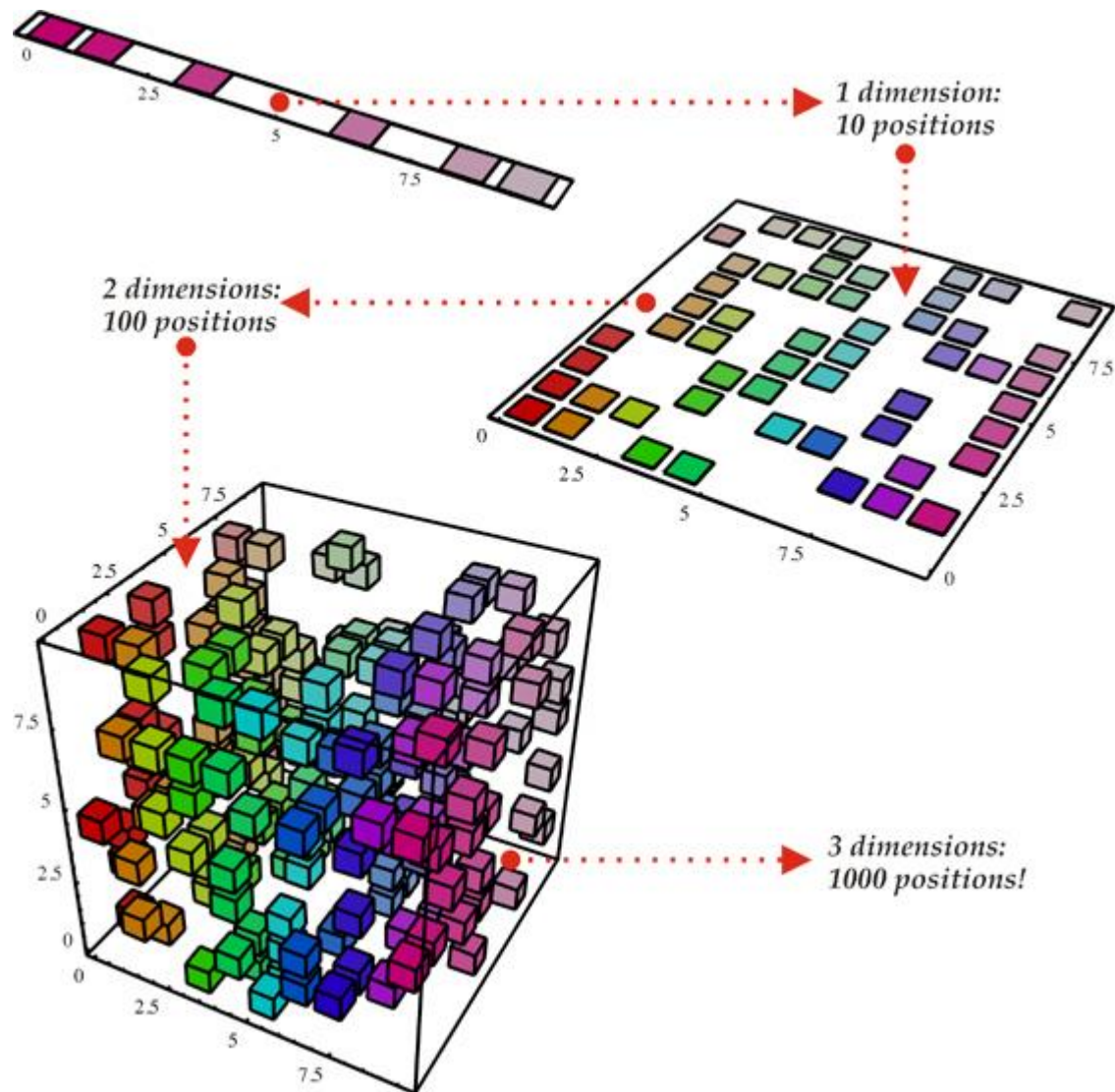
学习就需要从有限的例子中寻找到合理的 f

- 方向一：训练更多的数据：题海战术
- 方向二：加入先验知识：调整假设空间



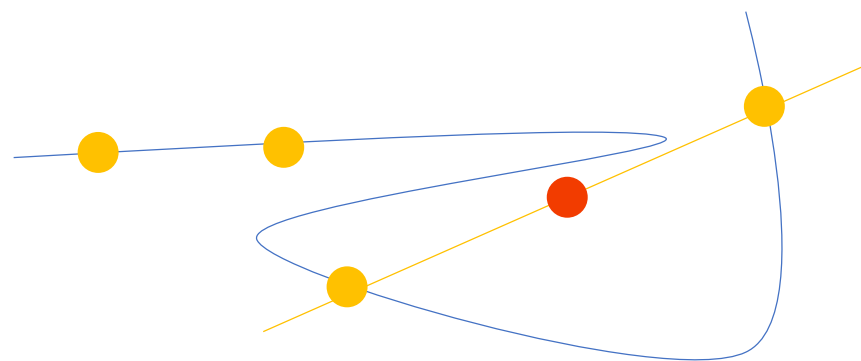
实际：无法被穷尽，
各式各样的变体。

维度诅咒



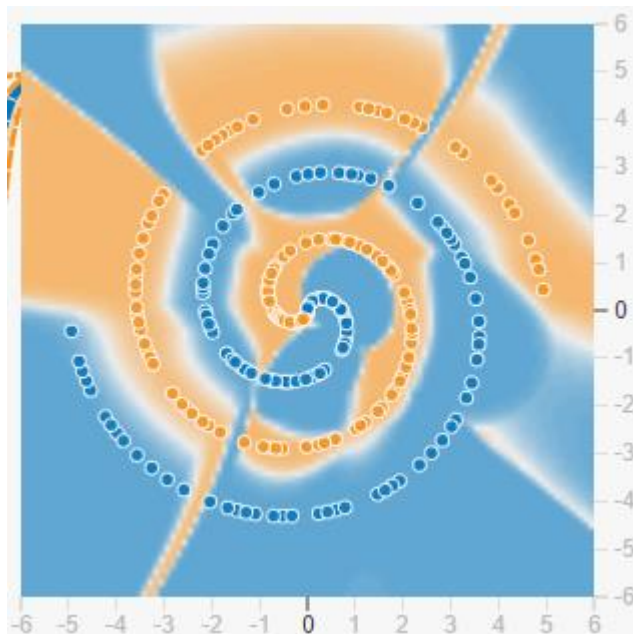
维度越大，我们越无法获得所有的情况。

面临没见过的情況，一般是將左右的情況**平均一下**。但是这种方法在高维数据下并不适用。

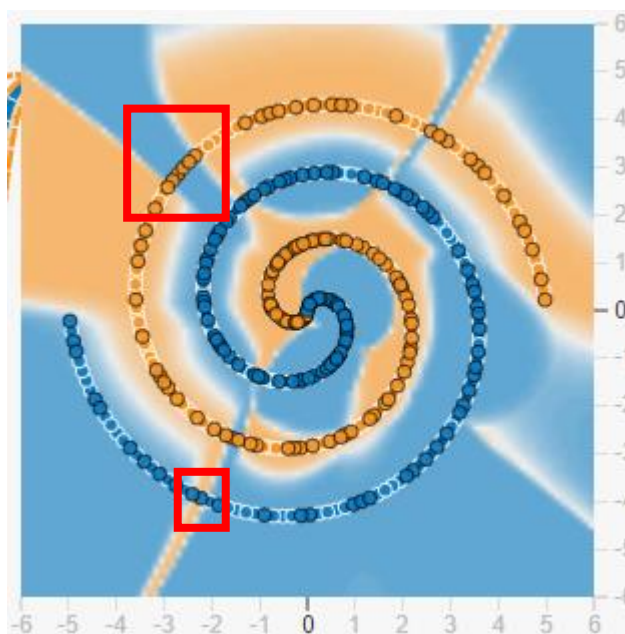


关于 f 的寻找

能够拟合训练集的关联 f 并非唯一



训练集



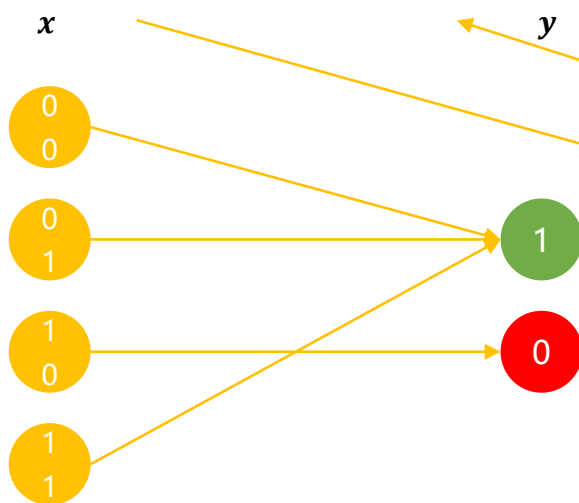
加入测试集（深色点）

光有大数据是不够的，难点在于测试集上的表现。

无法保证在训练集上表现良好的数据在测试集上也同样表现良好

无免费午餐

No Free Lunch Theorem

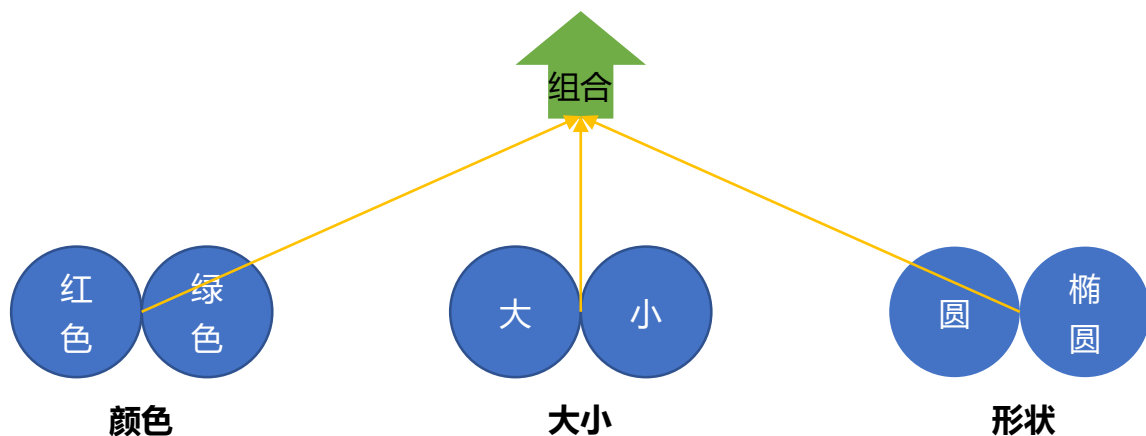


Any two optimization algorithms are equivalent when their performance is averaged across all possible problems



分布式表达

Distributed representation

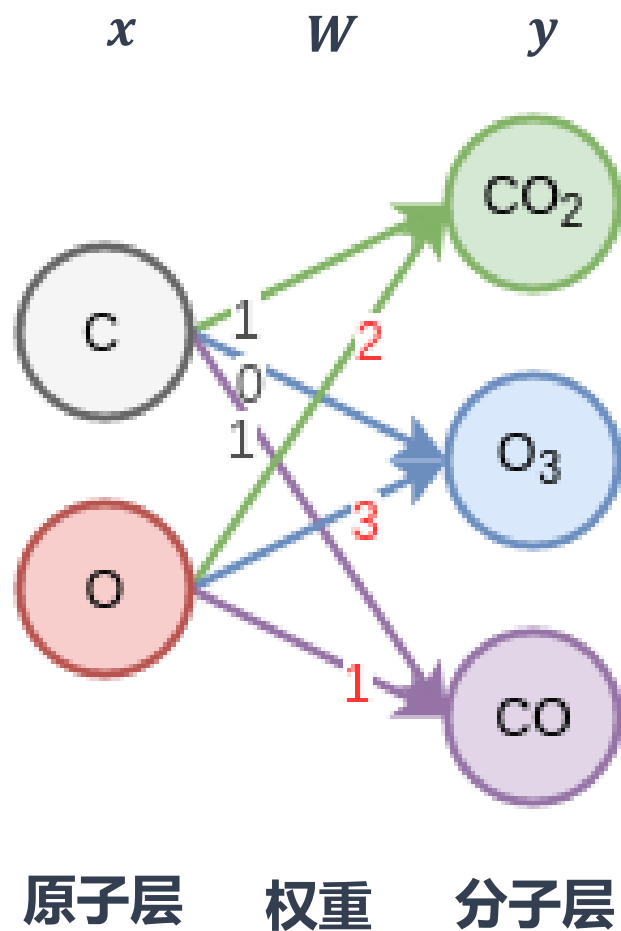


原本需要8个不同情况，现在只需要6个。
因为8个变体是又3种因素组合而成的。

distributed representation:
to disentangle the **factors** of **variation**

椭圆这个factor实际上也是有变体的，
可以以相同的思路继续拆分，继续降低
训练所需数据量。

神经网络



$$y = \text{relu}(Wx + b)$$

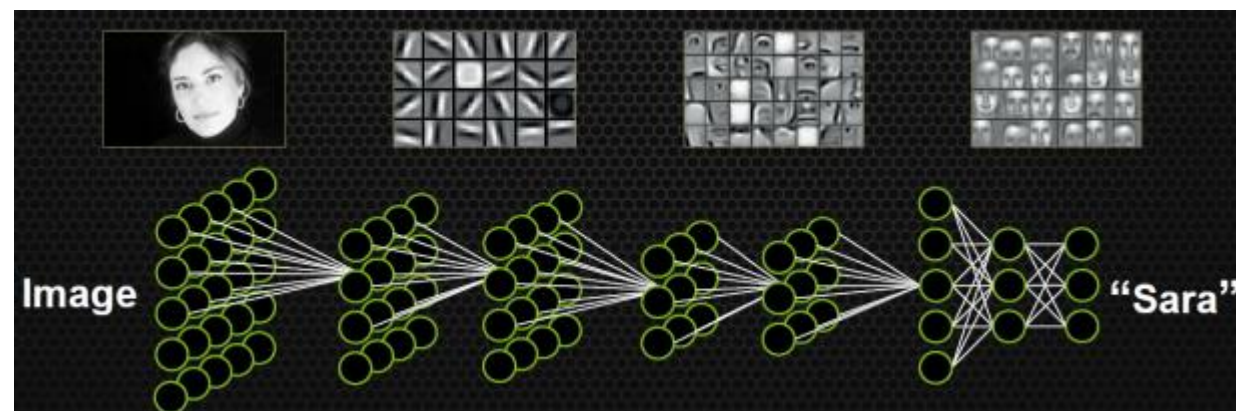
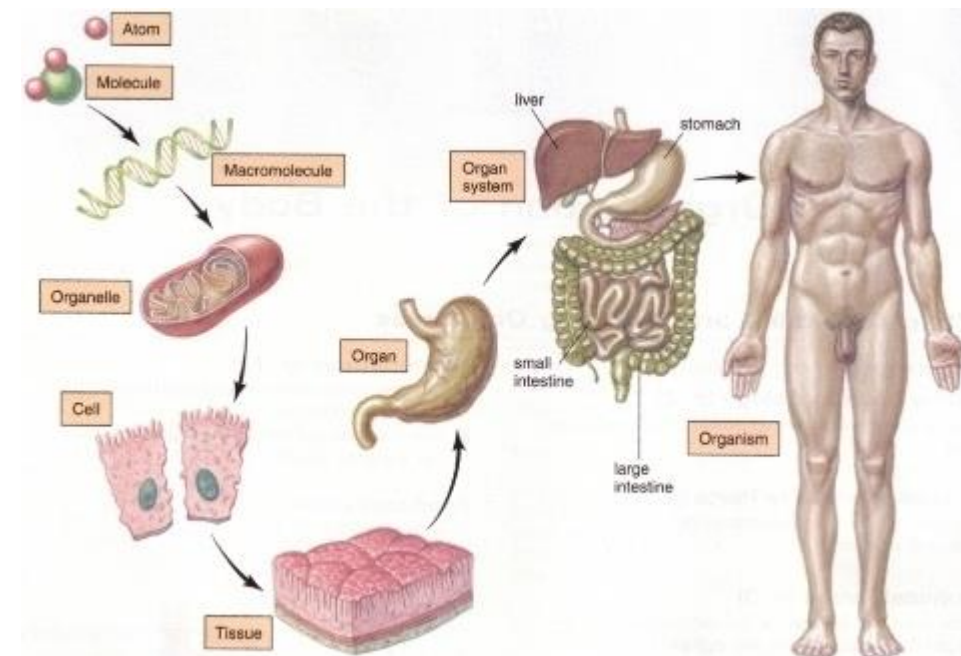
x : 输入

W : 控制拿多少个不同的原子

$\text{relu}()$: 非线性

b : 偏移, 超过该阈值才激活

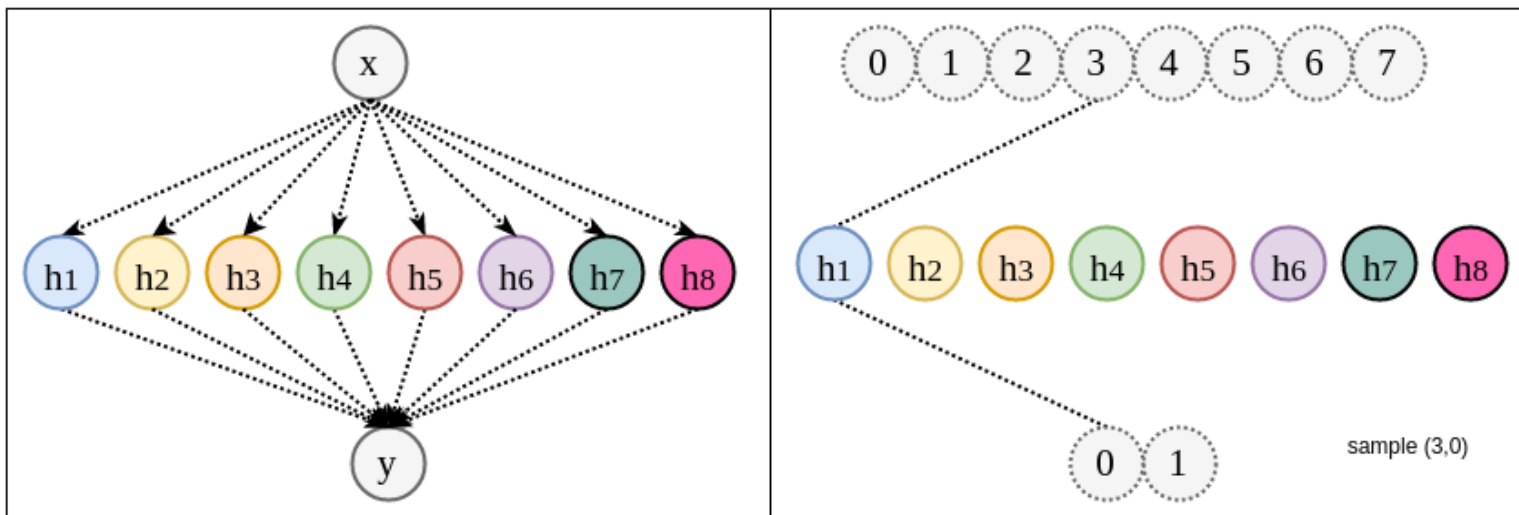
y : 输出



迭代变换

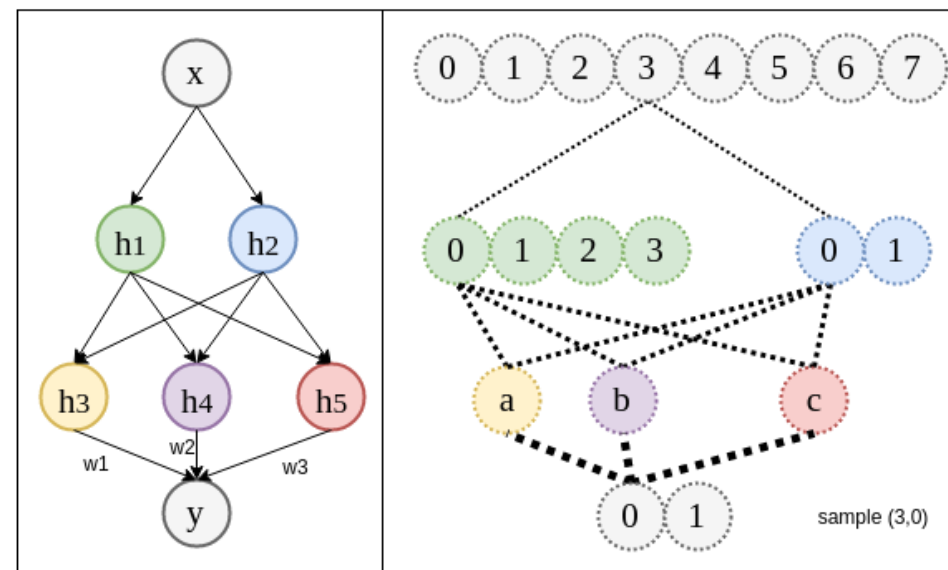
分布式表达是将变体拆分成因素。
但是如果用无限节点的浅层网络，所拆分的变体**并不会在不同样本之间形成共享**。

在浅层网络中只负责学习自己的关联，
而在深层网络中，那些**共用相同因素**的
样本也会被间接的训练到。



深层神经网络加入了
第二条先验知识：
迭代变换

关键：因素的共享

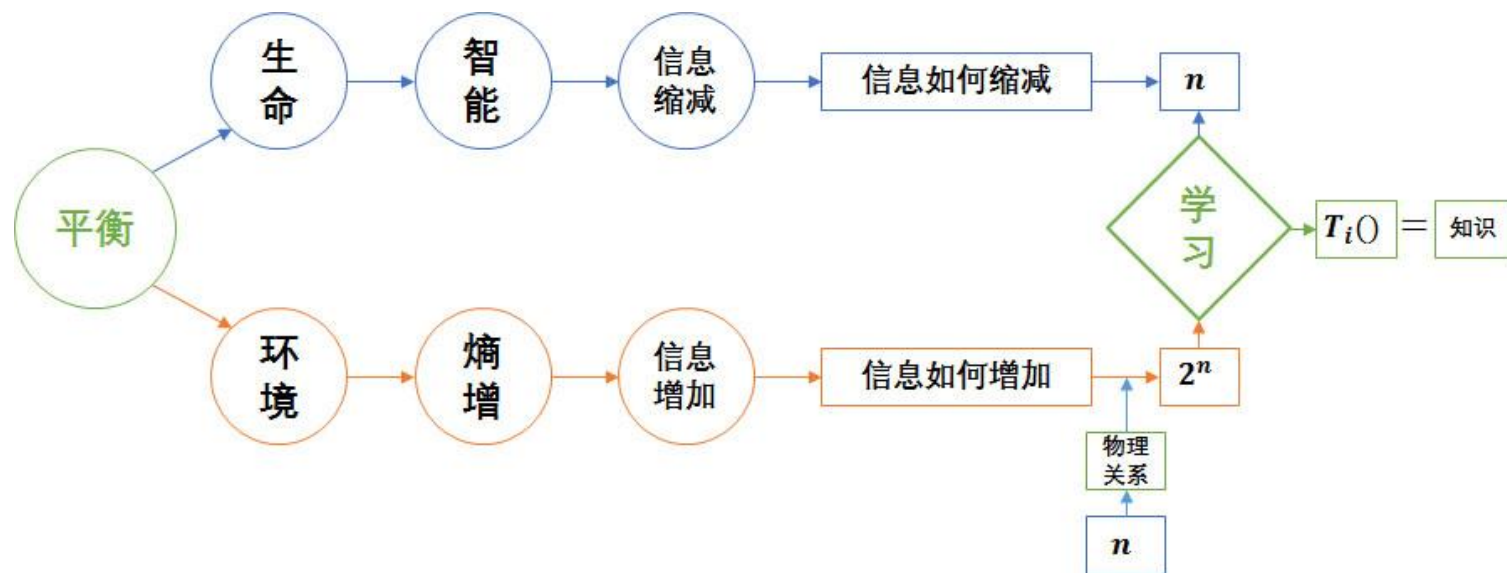


深层神经网络

学习的过程是因素间的关系的拆分，关系的拆分是信息的回卷，信息的回卷是变体的消除，变体的消除是不确定性的缩减。

自然界两个固有的先验知识：

- **并行**：新状态是由若干旧状态并行组合形成。
- **迭代**：新状态由已形成的状态再次迭代形成。



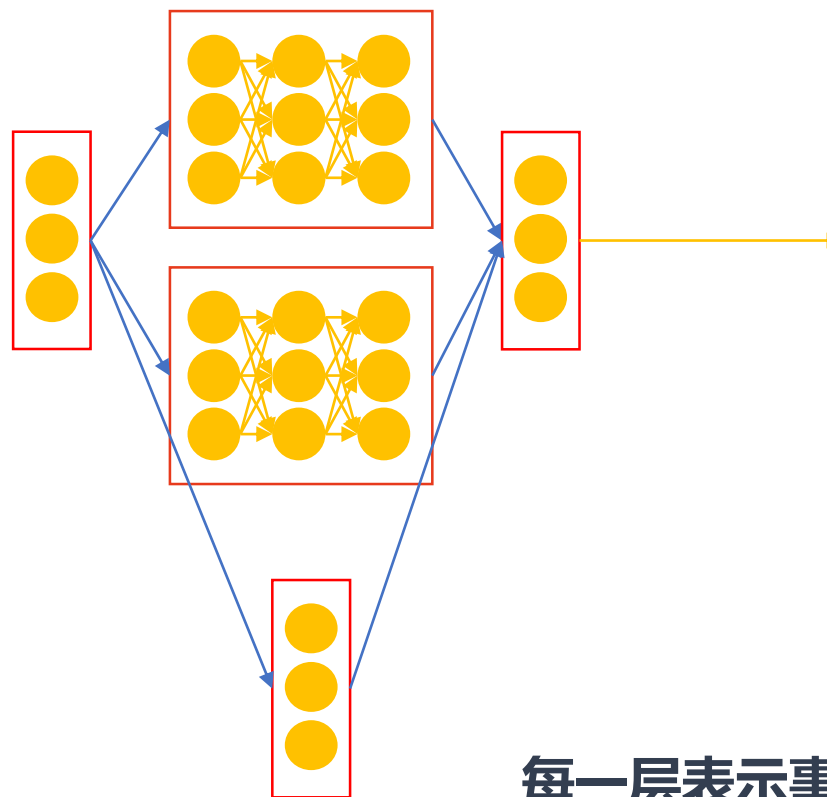
应用：设计理念

基本概念

- **拆分因素**：将变体拆分成因素，降低训练所需数据量。
- **因素共享**：使所拆分的因素在不同的样本之间形成共享，可以用等量的数据训练出更好的模型。

误区1：深度学习并非**万能**，应用先验知识的前提是数据可以以先验知识的方式生成

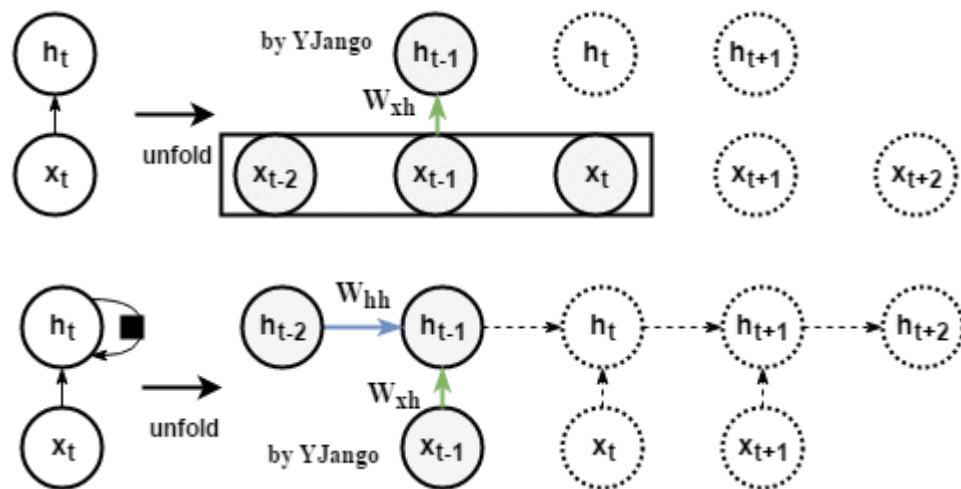
误区2：深度学习**没有固定形式**，加入不同先验知识，形成不同神经网络。



每一层表示事物的一种状态
设计神经网络时，要以“**层**”
为单元

神经网络变体

循环层：**时间共享**

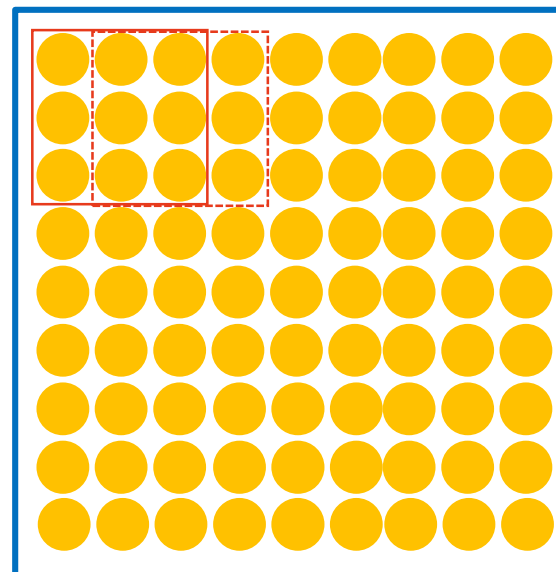


如果用前馈层，每个圆圈表示100个节点，那么前馈层处理时序相关性时就需要学习300个权重。

但如果知道不同权重在**时间下是共享的**，那么就只需要学习200个权重。



卷积层：**空间共享**



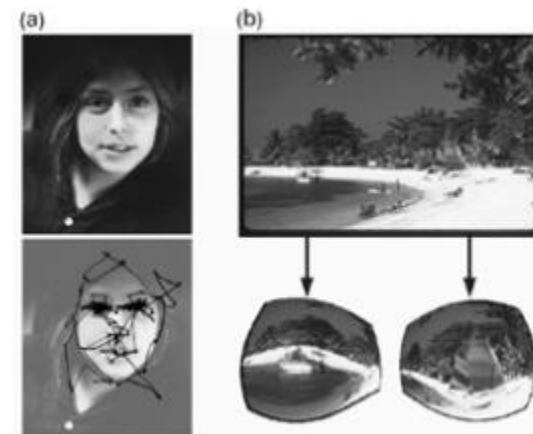
如果用前馈层，那么需要学习81个权重，但如果知道这些权重在**空间下是共享的**，那么可能只需要学习9个权重。原本一张图片在前馈层中只能用于学习一次权重，在卷基层中却可以学习很多次

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved Feature



调整假设空间

01

增加共享

降低数据量

02

增加惩罚

对那些不符合理念的范围进行惩罚

03

优化起点

优先从哪里找起

04

降低变体数

数据预处理

网络结构的设计

因素拆分、因素共享

网络学习的优化

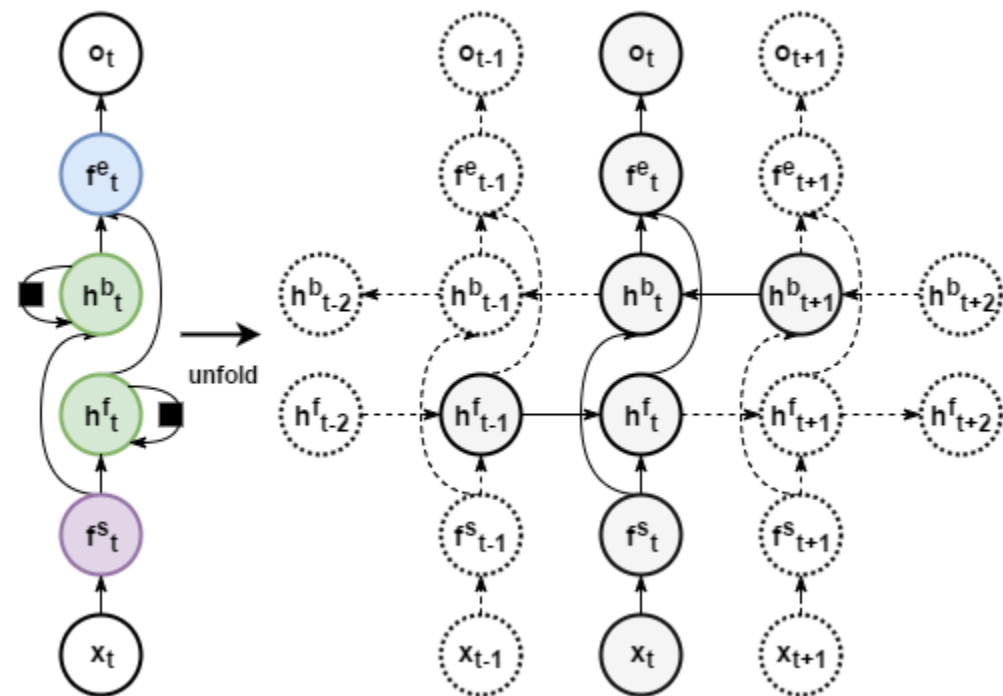
梯度下降所面临的问题、或寻找其他优化方式

设计自己的神经网络

神经网络其实并不黑箱，真正黑箱的是你的Task。

设计神经网络就是寻找在你手头的Task上利用因素拆分和因素共享的合理方式

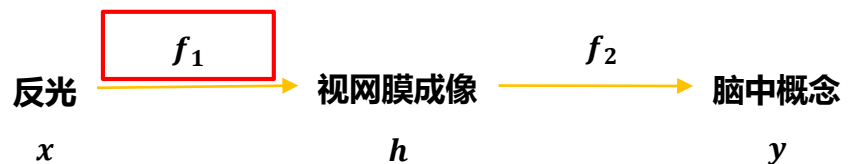
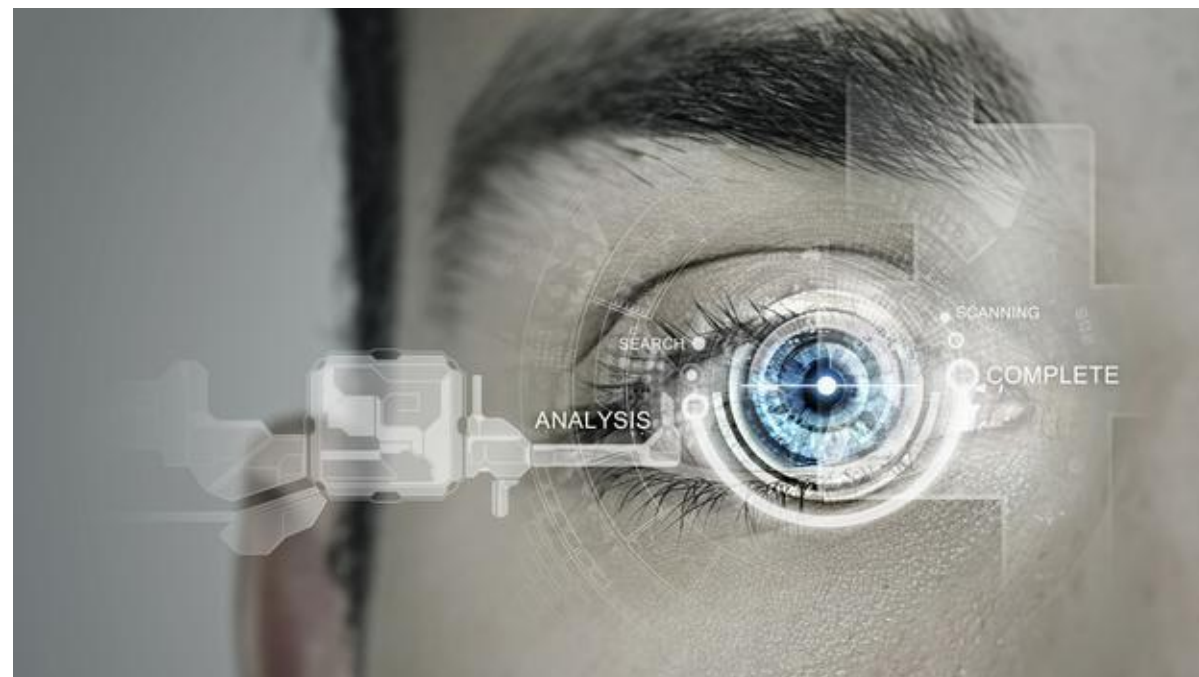
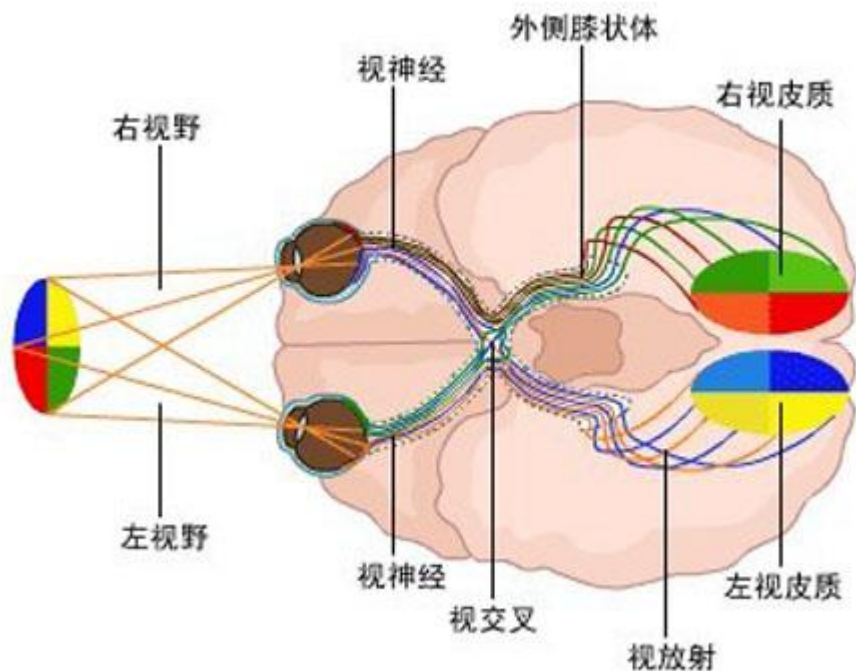
可以先经过前馈层再经过双向循环层再经过前馈层最终得到你的结果。



两个方向相加的 双向循环层一般比相并的效果好

迁移学习

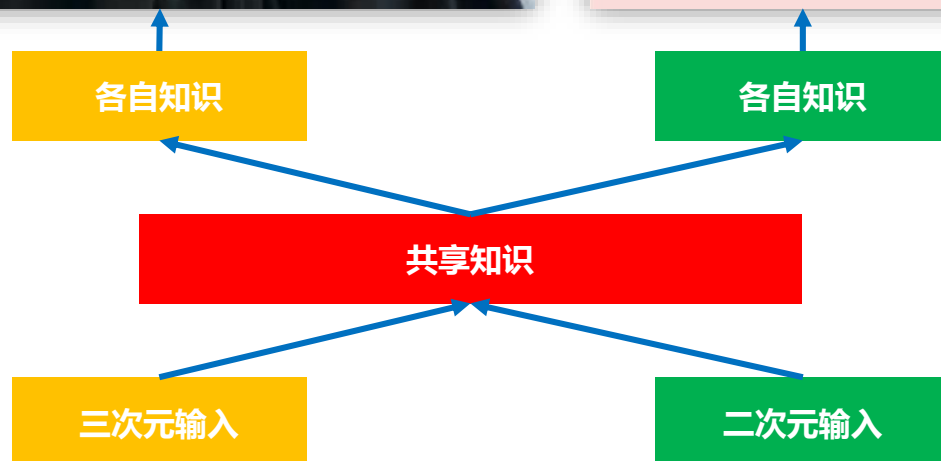
Transfer learning



利用**因素共享**这一特点
将一个任务中学习到的关联应用到其他任务当中去。

多任务学习

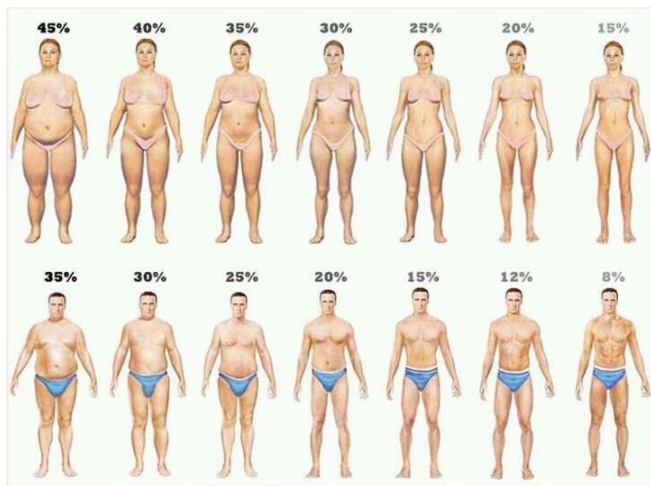
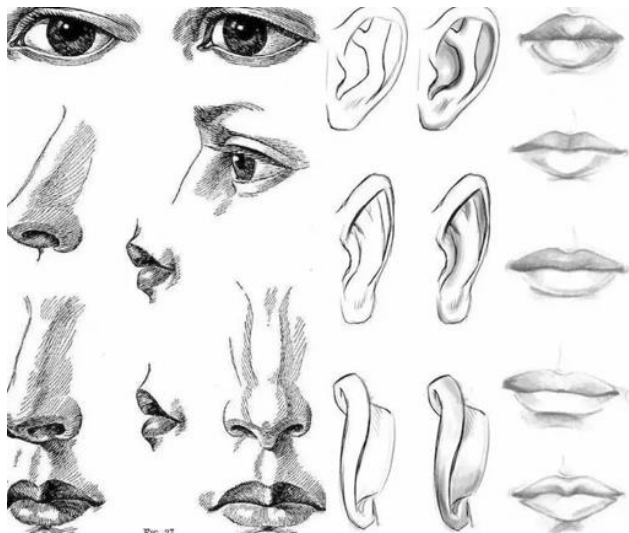
Multi task learning



先验知识：
与迁移学习一样是利用**因素共享**这一特点

作 用：
扩增额外训练数据
约束想要寻找的关联 f

跨层组合



先验知识：跨层组合

前馈神经网络是不允许跨层组合的

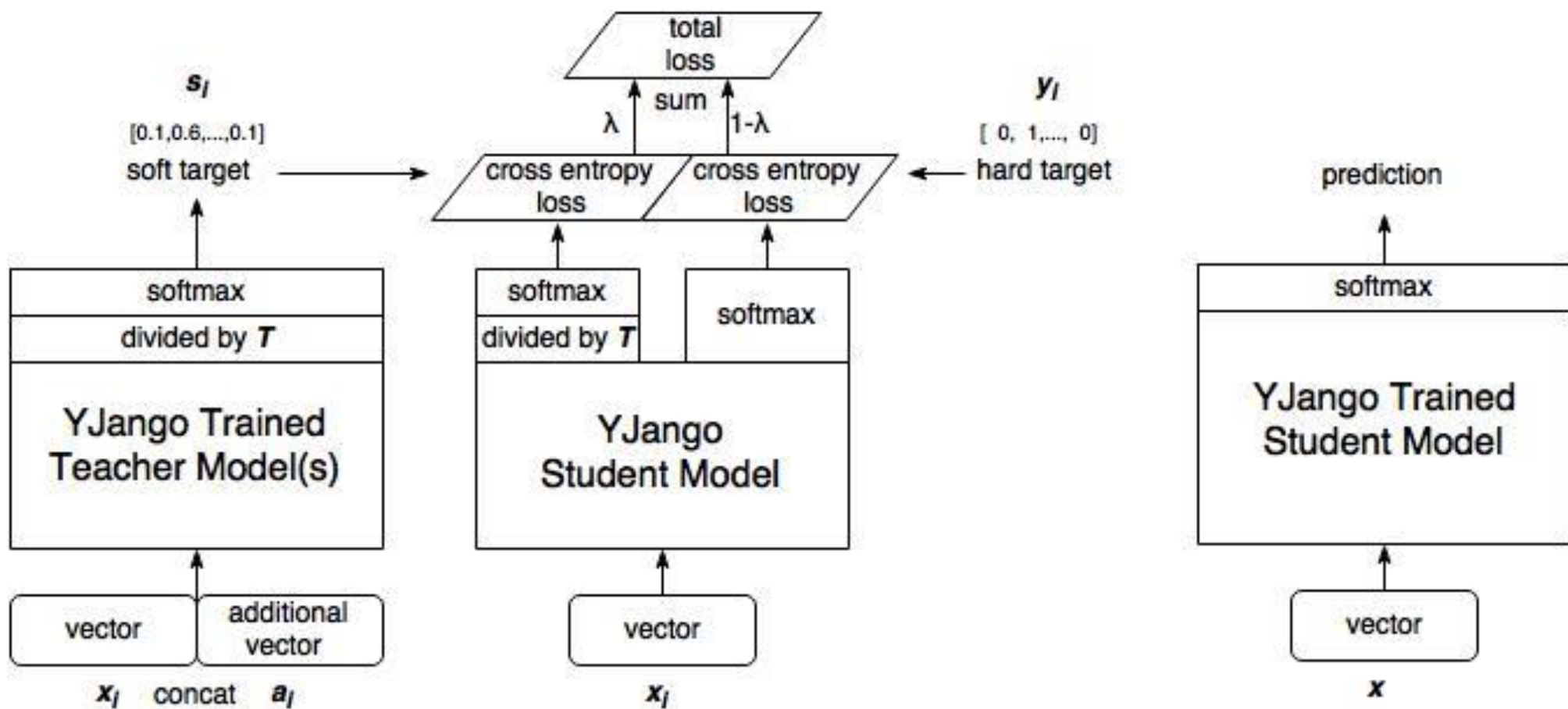
但现实中是否有跨层组合的现象？

代表网络：残差网络

眼睛 + 身形 + 着装 → 来判断一个人

蒸馏模型

Distillation

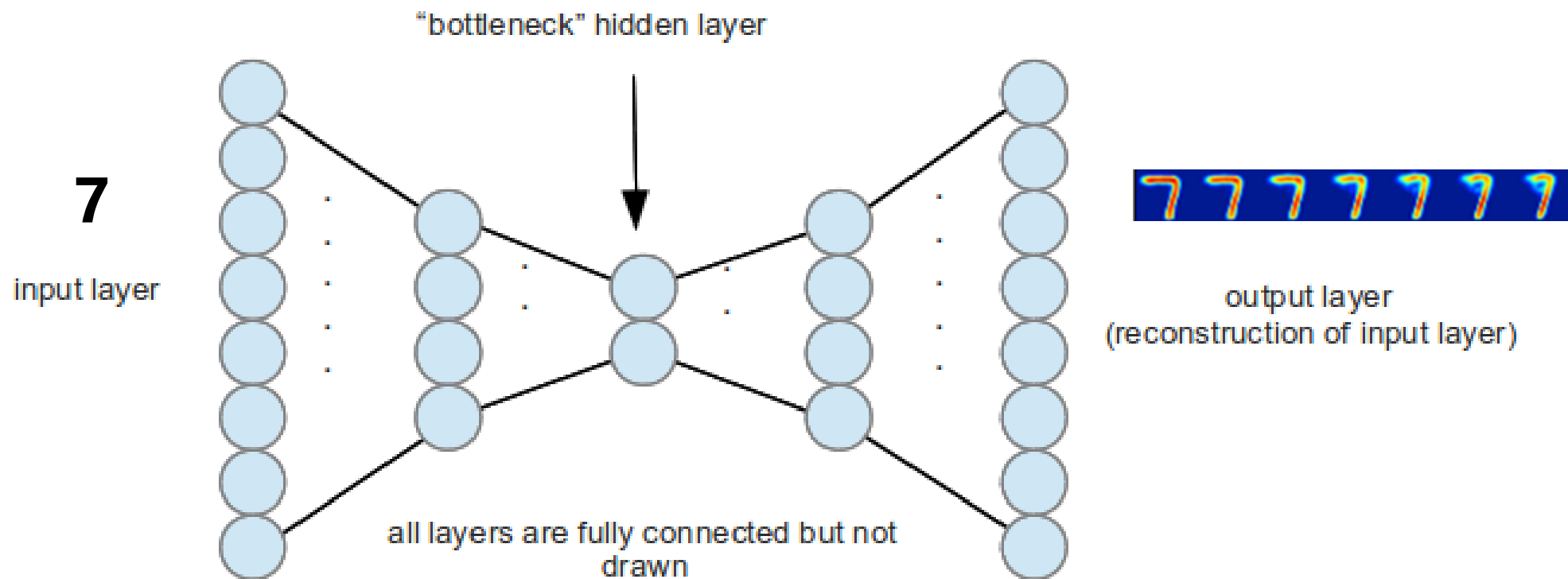


详细参考：[点我](#)

本质属于迁移学习，但使用的是不同的迁移方式

自动编码器

Autoencoder



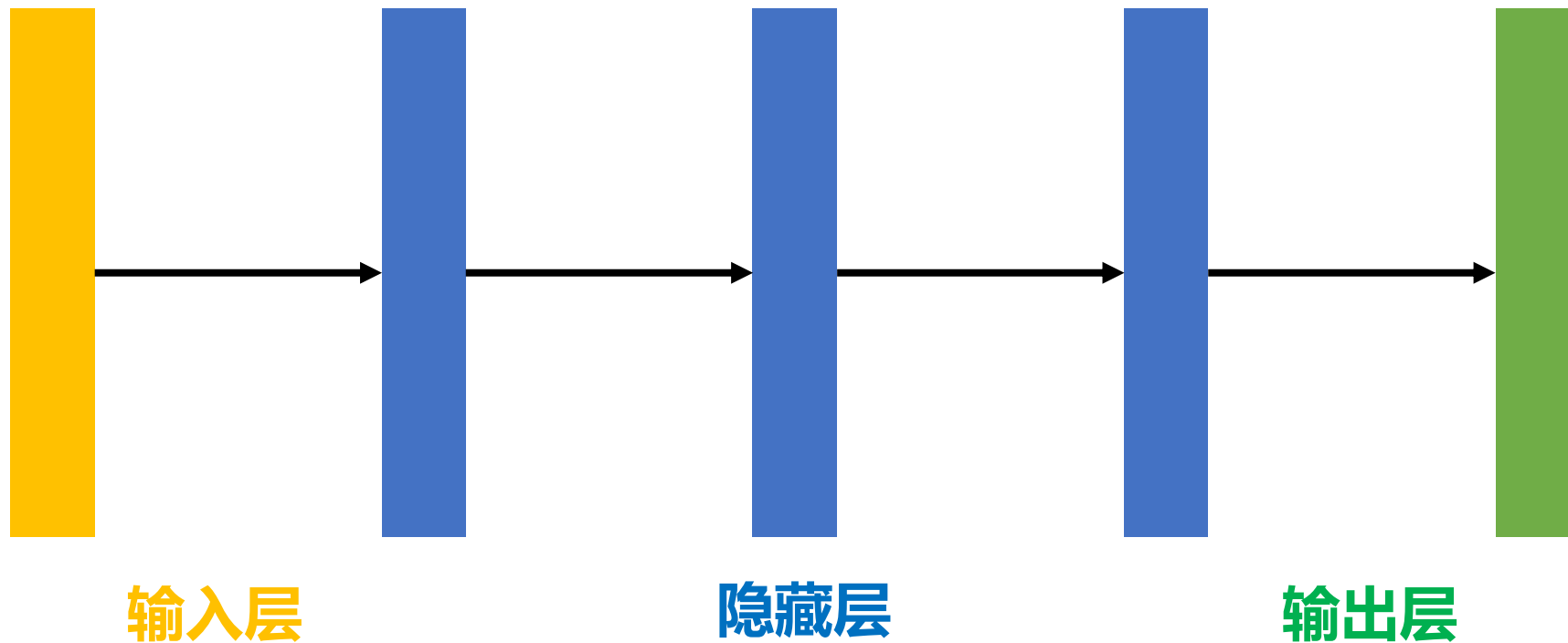
8个苹果 4个因素

自动编码器是对利用并行与迭代的这两个先验知识，来**操控变体**的一种技术。

To disentangle the factors of variation

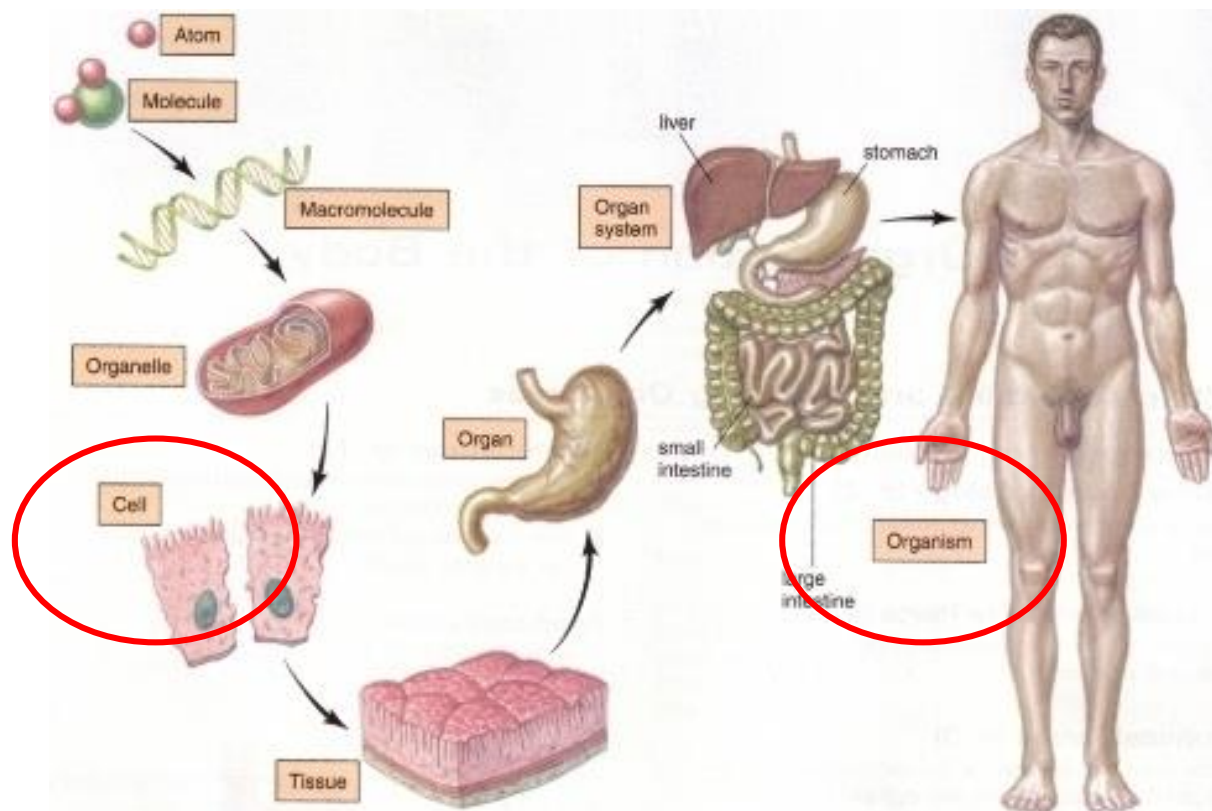
可用作生成模型，也可用作特征工程

Batch Normalization



在隐藏层上也来标准化从而降低变体数量，降低拟合难度

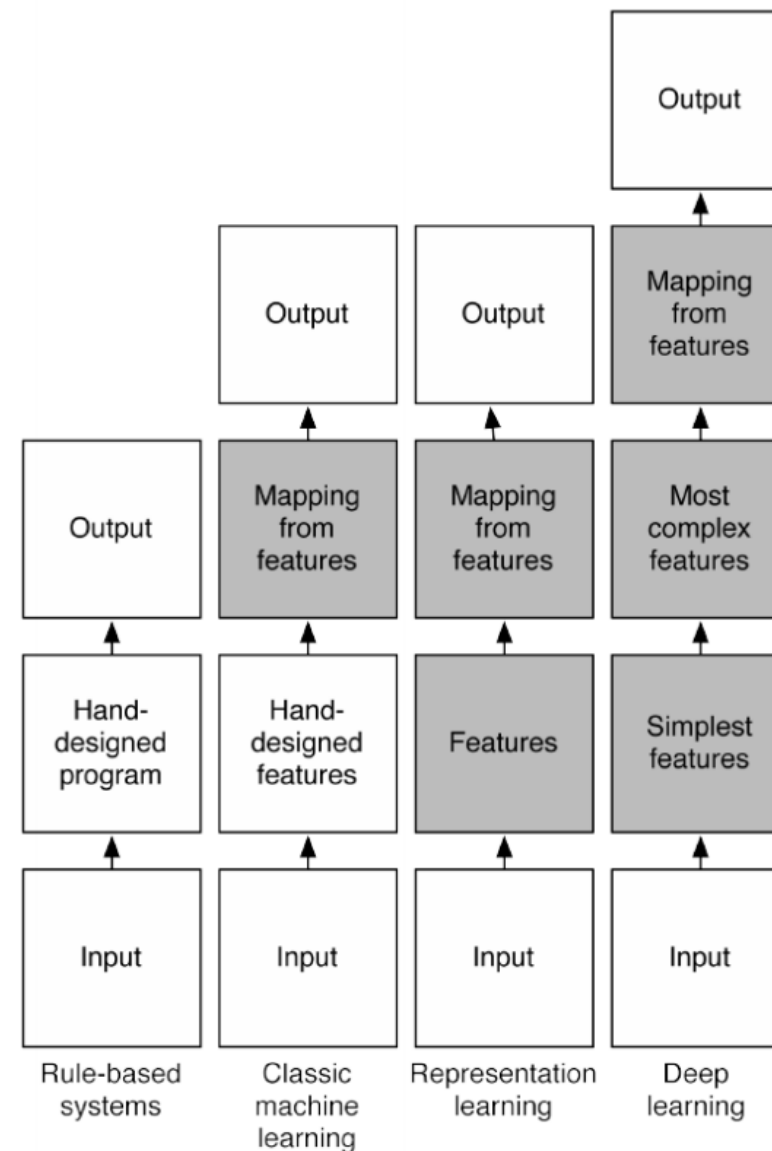
端到端



输入：细胞层

输出：系统层

End to end



抑制过拟合

Overfitting

dropout (遗忘) :

细节有时也能形成规律，但不会每次都形成。
遗忘可以去掉那些偶然形成的细节规律，提高普遍性。

shuffle (乱序) :

训练的样本不要有固定顺序，而要随机打乱，
同样可以抑制偶然形成的细节规律。
比如不要一直从abandon开始背单词一样。

L2 regularization (保持最简化) :

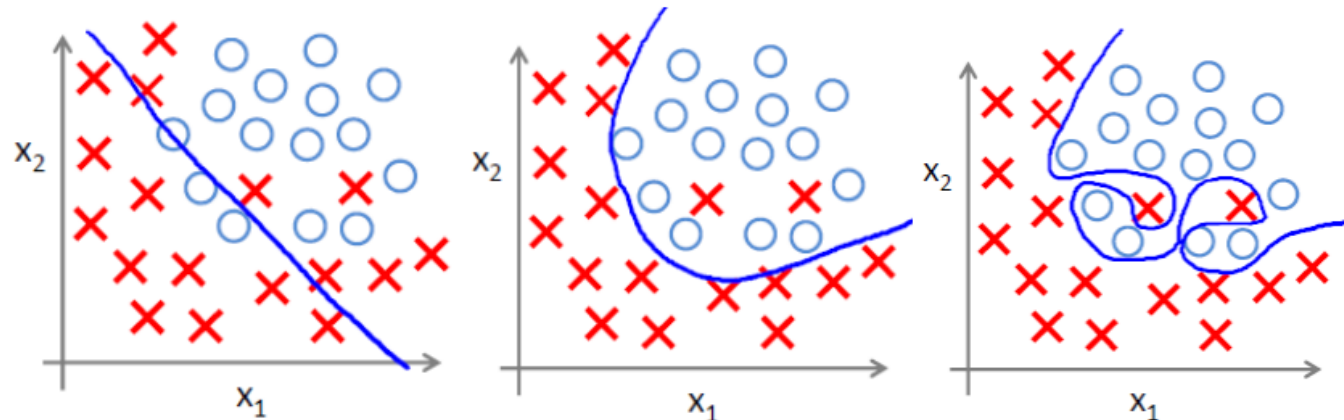
解决的方案不要过于复杂。
不然只能顾及特例而失去普遍性。

mini-batch (多题一起做) :

相互比较后得出结论。
比如同时看两本描述不同的书可以得到更好地理解。

noisy layer (加噪音) :

题目加入一些干扰项、改变考前环境、教室、平时状态等。



人工智能对我们的影响



帮助操作物体



帮助操作数据



帮助建立关联

机械解放了靠肉身操作物体的我们，计算机解放了靠大脑来计算数据我们。曾经我们是靠人脑来学习关联。**而目前的人工智能可以解放靠人脑学习关联的我们。**更好的预测和决策。

额外资料：《超智能体》

教程

01

Python矩阵操作

<http://cs231n.github.io/convolutional-networks/>

02

深度学习环境安装

<http://blog.csdn.net/u010751535/article/details/50806073>

03

教程三步（gitbook）

https://yjango.gitbooks.io/superorganism/content/tensorflowji_ben_yong_fa.html

https://yjango.gitbooks.io/superorganism/content/daima_yan_shi.html

https://yjango.gitbooks.io/superorganism/content/daima_yan_shi_2.html

<https://yjango.gitbooks.io/superorganism/content/%E4%B8%A3%E7%A0%81%E6%BC%94%E7%A4%BA%v3.html>

04

教程三步（看云）

<http://www.kancloud.cn/yjango/superorganism/265442>

<http://www.kancloud.cn/yjango/superorganism/265443>

<http://www.kancloud.cn/yjango/superorganism/265444>

<http://www.kancloud.cn/yjango/superorganism/265445>



Thank You

ありがとうございました。